

# How Well Do Elected Officials Represent Their General Populations?

December 10, 2020

Jared Dec, Brenna Giacchino, Aidan Jackson

## Overview

The purpose of this project was to determine how the demographics of elected officials compared to the voting age population which elects them. While there are elected officials at many different levels of government, this analysis focused on those who hold local positions at the town or city level. Local elections have been noted for both how “scale...can affect political participation”<sup>[2]</sup> as well as being “the nation’s foremost venues for the study of political behavior in the context of significant racial and ethnic diversity”.<sup>[1]</sup> Considering this, voters in local elections may be motivated by the greater influence their vote would have relative to the total voting population, or also by candidates who are from their community as opposed to being geographically distant. Therefore, local elections have the potential for more diverse candidates and the opportunity to be more reflective of the populations they represent.

## Research Questions

While the scope of this project will not be to determine any causal relationships between demographics and elected officials, it will investigate whether elected officials at the local level are representative of the voter populations they represent. We focused on the following research questions to determine this:

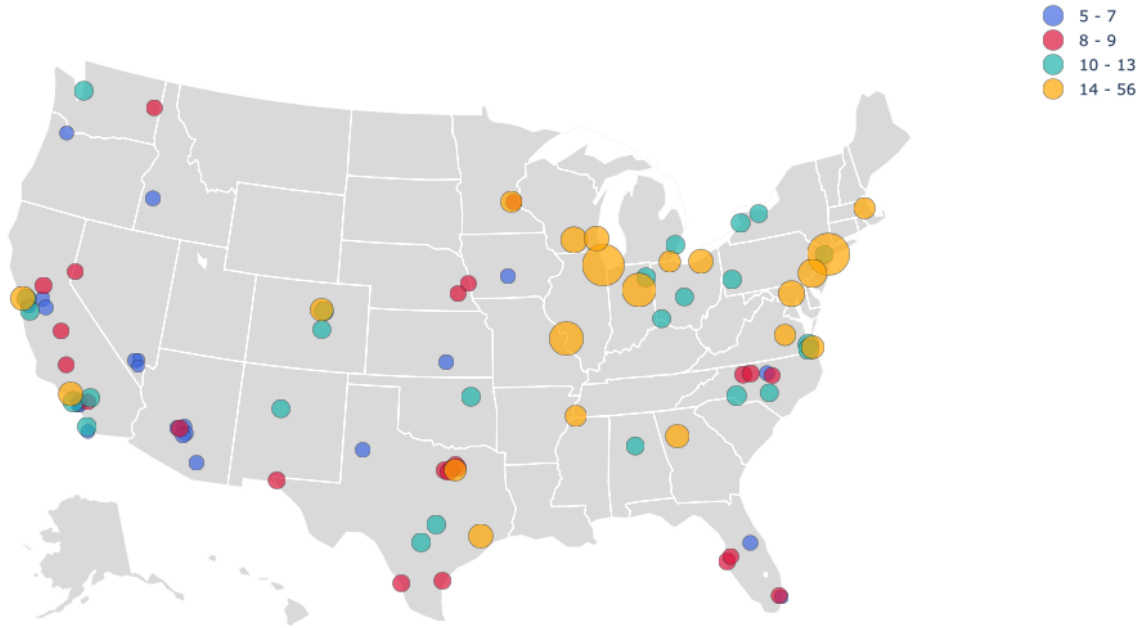
- How do the demographics of the voting age population in states in the U.S. compare to their elected officials?
- Statewide, are Males more represented than Females among their elected officials?
- Statewide, are Non-white males more represented than Non-white females among their elected officials?

## Data Sources

### *Elected Officials Data Set*

The primary [data set](#) was provided by the [Reflective Democracy Campaign](#), a group which investigates and promotes women and people of color in elected offices across the United States. The group collects data about demographics of elected officials from local to national governments, as well as conducts research to determine how systemic barriers may be overcome by these candidates. Analysis centered around a collection of **local officials in the 100 most populous cities in the U.S. that were in office in September of 2020**. The primary positions occupied by the local officials are city council members and mayors, although other elected positions such as City Attorney or City Clerk are included if applicable. The scope of the elected officials in the data set is restricted to either the city or county level.

**Figure I** shows the distribution of data points from the elected officials data set. Each circle represents a jurisdiction plotted at its geographic location on the map. The size of the circle is proportional to the number of data points, or elected officials, represented in the data set from that location, while the color categorizes the locations by quartile.



**Fig I.** 2020 Elected Officials Data Count by Jurisdiction

A big takeaway from this graph is that a majority (75%) of cities fall within the 5-13 data count range, a notably small sample size for large cities. To increase the degrees of freedom in the analysis, the data points were grouped by state. This shifted the analysis to how city- and county-wide demographics of elected officials compared to their state demographics. The distribution of data points by state is shown below in **Figure II**.

**Fig II.** Data Point Count by State in the Elected Officials Dataset

Although 33 states are represented in the elected officials data set, an arbitrary cutoff of states with 30 or more data points could be considered more reliable for producing results less prone to random chance. Thus, occasionally in the analysis states with greater than 30 data points may be treated as more principal to the results than others. In **Fig. II**, this begins with Pennsylvania.

### ***U.S. Bureau publication of 2018 voter-age demographics***

A supplemental data set is required in order to have information on the demographics of the voter population in each state for comparison with the data on the elected officials. This was obtained from a US Census Bureau [publication](#) on 2018 voter-age demographics. Of note is that this data set only estimates the population of voting-age US citizens for each state, and not the population of registered/eligible voters or those who were verified as having actually voted. According to one study from the Pew Research Center from 2014, approximately 21.4% of eligible voters were not actually registered to vote.<sup>[3]</sup> This data set is also two years older than the primary data set, which will require the assumption that demographics in the voting-age US citizen population of each state has not meaningfully changed in that time.

### **Assumptions and Data Cleaning**

When combining data from two sources, certain assumptions and filtering techniques were required in order to make statistical comparisons. The most broad was that the electorate of the local elected officials matched the average demographics for the state as a whole, so that a comparison could be made between the demographic distributions of the two. The elected officials data, however, was sampled from the 100 most populous cities in the US, where urban demographics would be different from rural demographics or the statewide average.

Additionally, certain demographic variables were transformed from the Census data set to match that of the elected officials dataset. The majority of these were related to the race variable, where only some Census designated categories matched those used in the elected officials data set. The biggest difference was related to how an individual's Hispanic or Latino identity was recorded. In the Census data, this variable was recorded separately from an individual's race, while the elected officials dataset treated it directly as an individual's racial identity. Therefore, when merging the Census data with the elected officials data, all Hispanic or Latino categories were aggregated to form a single "Hispanic or Latino" population for each state. The race associated with individuals who were Hispanic or Latino, such as "Hispanic or Latino and White" or "Hispanic or Latino and Black", were overwritten solely with the "Hispanic or Latino" category.

Finally, certain other categories were aggregated or dropped as required for comparison. For example, the "Asian" and "Native Hawaiian or Pacific Islander" categories were distinct within the Census data set, but within the elected officials' data set they were a single "Asian American or Pacific Islander" category. Other categories within the elected officials data set were also excluded due to not having an applicable match. These included the "Unknown" racial category, which excluded a small number of candidates from the analysis. For both data sets, however, the majority of data categories were able to be directly compared with one another.

## Statistical Methods

Analysis was performed via a chi-squared hypothesis test, which was used to determine how well a sample distribution of data represents a population's distribution of data<sup>[4]</sup>. Specifically, each chi-squared test described how the demographics of local elected officials in a state compared to the demographics of voters in that state. The chi squared test assumes samples are normally distributed, which was not investigated for this analysis. Alternative non-parametric methods, such as the Wilcoxon Ranked Sum test, could be used to avoid this assumption.

As previously mentioned, although the elected positions are local the analysis was conducted on a state by state basis to increase the degrees of freedom for each statistical test. The test proceeded under a null hypothesis that there is no difference between the demographics of the statewide voters and their local elected officials, with an alternative hypothesis that there is a difference (corresponding to a two-tailed test). We chose to define the metric for statistical significance as a p-value of 0.05, representing a 95% confidence that differences were not due to random chance.

## Results

- (1) *How do the demographics of the voting age population in states in the U.S. compare to their elected officials?*

**Fig III.** P-Values for Voting Age and Elected Official Population Demographics

In general, as shown in **Fig. III**, most states have a statistically significant difference in distribution and are colored in blue. In other words, racial and gender demographics of elected officials from a state are not representative of the state's actual population distribution. The exceptions are California, Texas, Arizona, Colorado, Idaho, New Mexico, and Virginia. While there is no absolute difference between different regions of the country, most of the states with non-significant differences are located in the Western Census Division, with the exception of

Texas and Virginia, which are both in the Southern Division. For this test, no states in the Midwest and Northeast divisions have non-significant differences.

However, not all states reflect the patterns that we initially anticipated. In Alabama, for instance, African-Americans are disproportionately represented among elected officials compared to their portion of the general statewide population. This could be because all the officials in our dataset for Alabama hold positions within the city of Birmingham, which according to the 2019 Census, has a population that is 70.5% African-American.<sup>[1]</sup> This is an example of how the limited scope of the elected officials data set caused irregularities. Another potentially contributing factor was that Alabama had only 10 data points total in the elected officials dataset.

**Fig IV.** Comparison of Race and Gender of Population and Elected Officials of Massachusetts

Another example of irregularities caused by the limited scope of the data set is Massachusetts, shown in **Fig. IV**, which has only 14 recorded officials. The population of Massachusetts is approximately 1.5% Multiracial Female, but there are two Multiracial Female elected officials in the data set making up 14.29% of the total.<sup>[2]</sup> Conversely, roughly 40% of Massachusetts's population is White Female, but there are also only two White Female elected officials in the data set, which is the same representation as Multiracial Females. This may be the source of statistical significance observed in this chi-squared test. Thus, while the results suggest a

narrative about the Western United States electing a more representative group of officials, the distribution is spread too thin for many of these states. There are effectively 12 potential race/gender categories for each state, and for many states, not many more elected officials than 12. Further research questions attempted to mitigate this problem by restricting the number of comparisons.

*(2) Are Males more represented than Females among elected officials?*

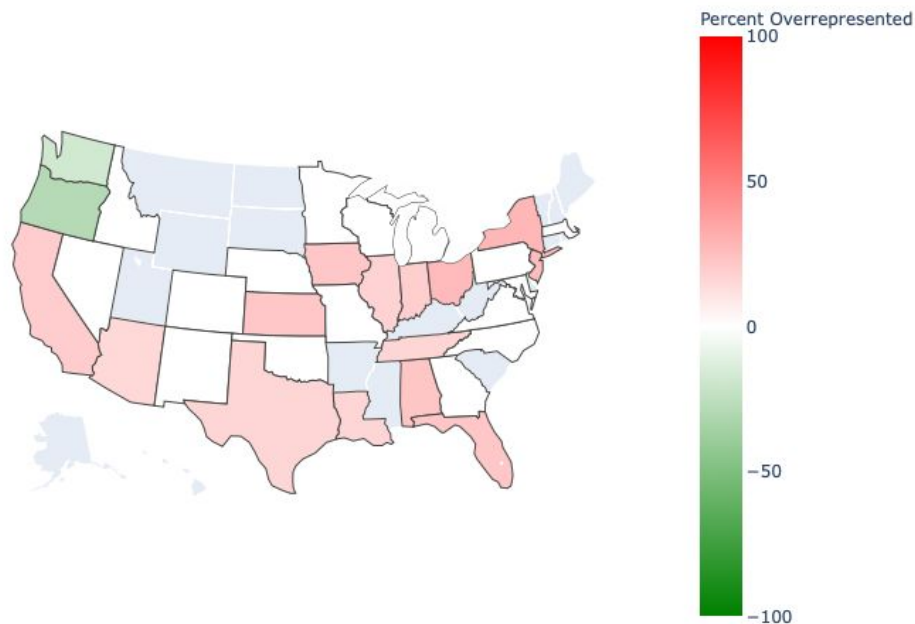
The distribution of gender among elected officials was also compared with the distribution of gender among the general population. The overall distribution of males and females across the 33 states in the data set is shown in **Fig. V** below.

**Fig V.** Elected Official Gender Distributions by State

Clearly, male elected officials are highly represented for most states. On the other hand, the Census data shows that the distribution of males and females for most states is between 45 and 55 percent in each direction, so it is safe to assume that the population distribution is close to 50% countrywide. **Fig VI** shows which states had significant differences between Male and Female distributions, in blue, and those which did not, in white.

**Figure VI.** P-Values for Gender Representation.

While **Fig VI** may show which states have statistically significant differences in gender distributions, it doesn't describe in which "direction" they differ or by what magnitude. That is, for the blue states, the figure doesn't describe whether males or females are more highly represented compared to the state's population. To answer this question, a second figure was produced that displays the percentage of the overrepresentation of Males among elected officials. For this analysis, "overrepresentation" was defined to be when a given demographic makes up a higher portion of the elected officials for their state than their distribution among the population, and which was determined from the test to be statistically significant.



**Fig VII.** Overrepresentation of Males.

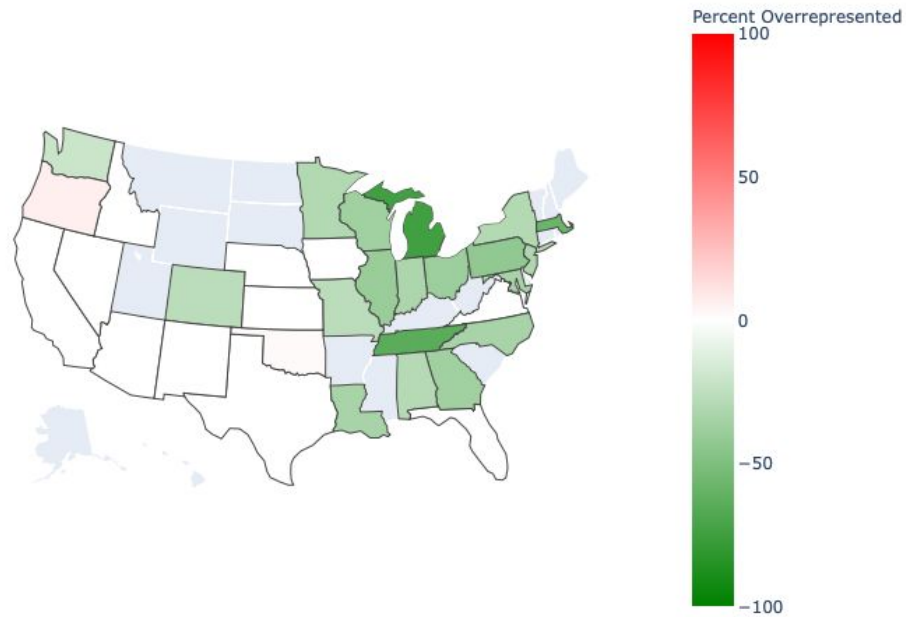
Shown in **Fig. VII**, the red states represent a higher percentage of Male elected officials, and the green states represent a higher percentage of Female elected officials. The white states are those in which there was no statistically significant difference between the gender distributions of elected officials and the state population, which match the above **Fig VI**.

Washington and Oregon are the only two states where there is an overrepresentation of Females among elected officials. Note that these two states are not represented by many data points, with Washington having 19 and Oregon having only five. Similar to the first research question, this lack of data could have contributed to irregularities in the outcomes for these states. On the other hand, it was surprising to see that California and New York had an overrepresentation of Male elected officials, by 19.5 and 24.8 percent, respectively. These states also had some of the highest number of data points and have results less due to random chance. Overall, out of 33 states, 15 showed an overrepresentation of male elected officials, two showed an overrepresentation of female elected officials, and 16 states showed no significant difference between male and female elected officials in reference to their populations.

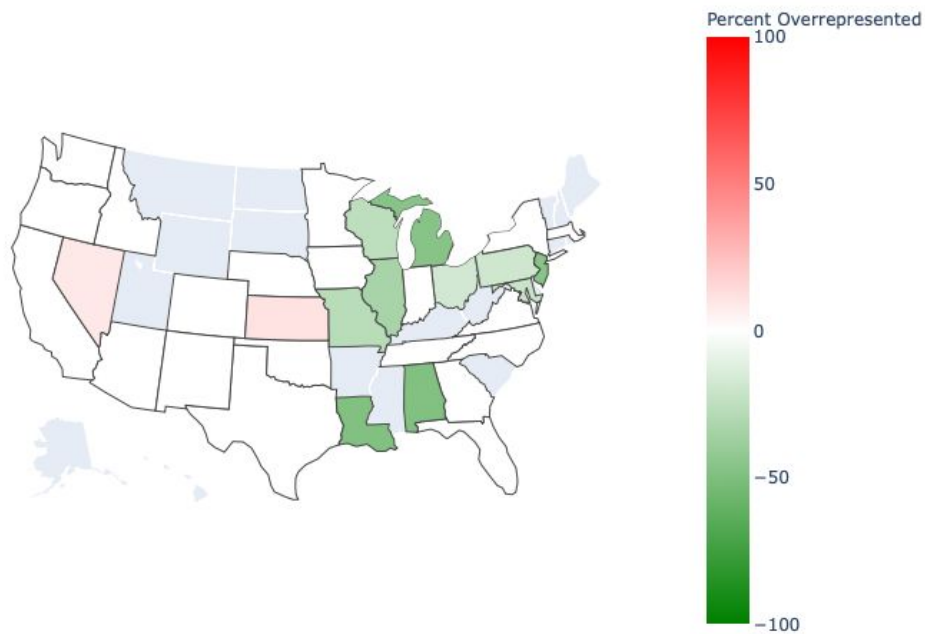
*(3) Are Non-White Males more represented than Non-white Females among elected officials?*

The final research question investigated whether representation of Non-white people differed between Females and Males. Unlike earlier research questions, this investigation proceeded in two phases. In the first phase, the distribution of the population and elected officials across race was examined only for Females in each state. This was performed under the null hypothesis that the distribution of Females by race in the population would have the same distribution as those in elected positions. The second phase repeated this examination for Males under the same conditions and null hypothesis. If the null hypothesis was rejected for a state, then the percentage difference between White and Non-white representation was calculated for either Females or Males as applicable. If the null hypothesis was accepted, then the percentage difference was assumed to be zero. Finally, the difference between White and Non-White representation was compared between Females and Males in each state to determine how representation varied.





**Fig VIII.** Overrepresentation of White Females versus Non-white Females.



**Fig IX.** Overrepresentation of White Males versus Non-white Males.

Shown in the **Figures VIII** and **IX** above, generally both Non-white Males and Females had a greater proportional representation than their White counterparts in the analysis. However, common studies of racial disparities in the US note that White people, and Males in particular, are overrepresented across government.<sup>[8]</sup> Therefore, examining this result in the context of the analysis likely highlights the failure of one of the key assumptions when combining the elected official and voter population data sets. That assumption was that urban elected officials, from the 100 most populous cities in the primary data set, would represent an electorate that is equivalent to the statewide average demographics. As noted previously, urban demographics

are primarily less White than suburban or rural areas, meaning that a statewide population usually contains more White potential voters than the urban populations that actually elected the officials in our data set.

An example of this is in Michigan, where the primary dataset contained solely elected officials from Detroit. While Michigan is on average 79% White, elected officials in the primary dataset representing Detroit were 82% Non-White. Although this was a significant difference in the context of the test, it is likely explained by the fact that Detroit is 85% Non-white,<sup>[9]</sup> which is approximately what is represented among the recorded elected officials. This led to a rejection of the null hypothesis for the test across both Males and Females for that state, which was a trend shown for many states in the above figures. However, given that these differences exist which were not accounted for by the choice of datasets, they are likely not reflective of the true population of elected officials and voter demographics in the US.

## Potential Biases

The main bias of the analysis that has been discussed throughout this report is not having sufficient data from the elected officials data set to represent an entire state's population demographics. In the preliminary stages of the project, it was decided to group the elected officials data points by state rather than by city for the following reasons:

- Individual cities often did not have enough data points to produce reliable chi-squared results; grouping by state increased the degrees of freedom for each comparison.
- The 2018 Census data set used to represent the voting population did not have resolution down to the city level, only by state

However, this state categorization for the chi-squared comparisons led to inconclusive results. A big concern, especially with the third research question, is that comparing an entire state's population distribution with that of one or several cities' may not be a robust comparison, even if the cities are large compared to others in the state. Other data collection organizations have noted that urban demographics have unique differences compared with states as a whole.<sup>[5]</sup> Therefore, in comparing urban elected officials with statewide population distributions, this analysis makes the assumption that these differences do not exist under the null hypothesis.

Another concern is that the elected officials data set only includes elected officials from the 100 most populous cities in the country. This means that states with higher populations are more likely to be on this list multiple times, so that these states will be better analyzed than others that have only one or no cities in the data set. Potential ways to address this would also include better refining the data sources used so that countywide officials may be included uniformly from across the US.

The final concern is that the elected official positions in the elected officials data are not consistent between cities. For example, some cities such as Anaheim and Cincinnati have only city council members and a mayor listed, while others like San Francisco have additional positions like Public Defender, Supervisor, and Attorney. This variation in the set of elected official members from each city could obscure results due to potentially missing data points.

Additionally, if the elected officials data set did contain consistent elected official positions for each city, it may have been more feasible to compare this with the city's population distribution, rather than the state's. Again, this would require finding an **alternative data set** that categorizes population demographics by city, rather than state.

## Conclusions

In general, the three research questions have generated three different conclusions. The first is that when broken into every combination of race and gender in the dataset, twelve categories stretch the elected officials' data too thin. Most states don't have many more data points than 12, meaning that for the race-gender test, results are inconclusive. For race alone, the comparison is effectively apples and oranges, city-level elected officials and state-level populations, which cause the results to come out skewed. This is because most cities in the dataset tend to have a higher percentage of Non-white people in their population compared to the state as a whole. For instance, Tennessee and Michigan both come out as over-representing Non-white elected officials, but both states only have elected officials for the cities of Memphis and Detroit respectively. These cities are both overwhelmingly Non-white in their general population which can explain this overrepresentation.<sup>[9]</sup>

However, it can be conclusively determined that Males are, as a whole, far more likely to be overrepresented than Females. As the portion of Male/Female populations should be roughly equivalent in most cities and states as a whole, the fact that half the states show an overrepresentation of Males does indicate that this is a significant result nationwide. It is also important to note that even traditionally liberal states like California and New York are significant in this relationship towards Males while still also having the greatest number of data points. Moreover, only two states have disproportionately more Female elected officials (Oregon and Washington). Still, the integrity of this finding is questionable, as both states have relatively few elected official data points in the data set.

To come to a more robust conclusion, one of two avenues would have to be pursued. The first option would be to not use state-wide populations but rather the underlying city-level populations. This would allow the elected officials to be compared with their direct electorate rather than assuming the electorate matches state level demographics. The second would be to use state-wide elected officials rather than city-level elected officials, and continue using state level electorate demographics. Each has its own issues, however. There may be not enough state-level officials to have enough degrees of freedom for a reliable chi-squared test. Without aggregating the city-wide elected officials, the degrees of freedom may also drop too low to produce reliable test results as well. For either option, it would be worthwhile to pursue how a new dataset would affect the results.

**Works Cited:**

[1] "U.S. Census Bureau QuickFacts: Birmingham, Alabama." *United States Census Bureau*, [www.Census.gov/quickfacts/fact/table/birminghamcityalabama/IPE120219](http://www.Census.gov/quickfacts/fact/table/birminghamcityalabama/IPE120219).

[2] "U.S. Census Bureau QuickFacts: Massachusetts." *United States Census Bureau*, [www.Census.gov/quickfacts/MA](http://www.Census.gov/quickfacts/MA).

[3] Kaufmann, Karen M. & Rodriguez, Antonio. Political Behavior in the Context of Racial Diversity: The Case for Studying Local Politics. University of Maryland.  
<https://cpb-us-w2.wpmucdn.com/sites.uwm.edu/dist/3/247/files/2016/07/SYMPOSIUM-Marschall-Shah-Ruhil-1520gbn.pdf>

[4] Lewis, Paul G. Size and Local Democracy: Scale Effects in City Politics. Arizona State University.  
<https://cpb-us-w2.wpmucdn.com/sites.uwm.edu/dist/3/247/files/2016/07/SYMPOSIUM-Marschall-Shah-Ruhil-1520gbn.pdf>

[5] Why Are Millions of Citizens Not Registered to Vote?, The Pew Charitable Trusts, 21 June 2017,  
[www.pewtrusts.org/en/research-and-analysis/issue-briefs/2017/06/why-are-millions-of-citizens-not-registered-to-vote](http://www.pewtrusts.org/en/research-and-analysis/issue-briefs/2017/06/why-are-millions-of-citizens-not-registered-to-vote).

[6] <https://www.statisticshowto.com/goodness-of-fit-test/>

[7] <https://www.pewsocialtrends.org/2018/05/22/demographic-and-economic-trends-in-urban-suburban-and-rural-communities/>

[8] <http://amsterdamnews.com/news/2017/nov/09/study-white-men-are-overrepresented-politics/>

[9] <https://www.census.gov/quickfacts/detroitcitymichigan>